

SEQ ID NO:1

SEQUENCE TYPE:Nucleotide

SEQUENCE LENGTH:21 BASES

STRANDEDNESS:single

TOPOLOGY:linear

MOLECULE TYPE:synthetic DNA

ORIGINAL SOURCE ORGANISM:bacteriophage lambda gt11

IMMEDIATE EXPERIMENTAL SOURCE:Oligonucleotide synthesiser; oligo d19

FEATURES:

from 1 to 21 bases homologous to upstream portion of lacZ gene  
flanking the EcoR1 site in bacteriophage lambda gt11

PROPERTIES:primes DNA synthesis from the phage vector into cDNA  
inserted at the EcoR1 site.

GGTGGCGACG ACTCCTGGAG C

21

SEQ ID NO:2

SEQUENCE TYPE:Nucleotide

SEQUENCE LENGTH:21 BASES

STRANDEDNESS:single

TOPOLOGY:linear

MOLECULE TYPE:synthetic DNA

ORIGINAL SOURCE ORGANISM:bacteriophage lambda gt11

IMMEDIATE EXPERIMENTAL SOURCE:Oligonucleotide synthesiser; oligo d20

FEATURES:

from 1 to 21 bases homologous to downstream portion of lacZ gene  
flanking the EcoR1 site in bacteriophage lambda gt11

PROPERTIES:primes DNA synthesis from the phage vector into cDNA  
inserted at the EcoR1 site.

TTGACACCAG ACCAACTGGT A

21

SEQ ID NO:3

SEQUENCE TYPE:Nucleotide with corresponding protein

SEQUENCE LENGTH: 1770 BASE PAIRS

STRANDEDNESS:single

## TOPOLOGY: linear

MOLECULE TYPE:cDNA to genomic RNA

ORIGINAL SOURCE ORGANISM:human; serum infectious for  
post-transfusional non-A, non-B hepatitis  
IMMEDIATE EXPERIMENTAL SOURCE:clone JG2 from cDNA library in lambda  
gt11

## **FEATURES:**

from 1 to 1770 bp portion of the PT-NANBH polyprotein

PROPERTIES: probably encodes viral non-structural proteins

CAA AAT GAC TTC CCA GAC GCT GAC CTC ATC GAG GCC AAC CTC CTG TGG 48  
 Gln Asn Asp Phe Pro Asp Ala Asp Leu Ile Glu Ala Asn Leu Leu Trp  
                   5                  10                  15

CGG CAT GAG ATG GGC GGG GAC ATT ACC CGC GTG GAG TCA GAG AAC AAG 96  
 Arg His Glu Met Gly Gly Asp Ile Thr Arg Val Glu Ser Glu Asn Lys  
                  20                 25                 30

GTA GTA ATC CTG GAC TCT TTC GAC CCG CTC CGA GCC GAG GAG GAT GAG 144  
 Val Val Ile Leu Asp Ser Phe Asp Pro Leu Arg Ala Glu Glu Asp Glu  
35                  40                  45

CGG GAA GTG TCC GTC CCG GCC GAG ATC CTG CGG AAA TCC AAG AAA TTC 192  
 Arg Glu Val Ser Val Pro Ala Glu Ile Leu Arg Lys Ser Lys Lys Phe  
 50 55 60

CCA CCA GCG ATG CCC GCA TGG GCA CGC CCG GAT TAC AAC CCT CCG CTG	240
Pro Pro Ala Met Pro Ala Trp Ala Arg Pro Asp Tyr Asn Pro Pro Leu	
65 70 75 80	
CTG GAG TCC TGG AAG GCC CCG GAC TAC GTC CCT CCA GTG GTA CAT GGG	288
Leu Glu Ser Trp Lys Ala Pro Asp Tyr Val Pro Pro Val Val His Gly	
85 90 95	
TGC CCA CTG CCA CCT ACT AAG ACC CCT CCT ATA CCA CCT CCA CGG AGA	336
Cys Pro Leu Pro Pro Thr Lys Thr Pro Pro Ile Pro Pro Pro Arg Arg	
100 105 110	
AAG AGG ACA GTT GTT CTG ACA GAA TCC ACC GTG TCT TCT GCC CTG GCG	384
Lys Arg Thr Val Val Leu Thr Glu Ser Thr Val Ser Ser Ala Leu Ala	
115 120 125	
GAG CTT GCC ACA AAG GCT TTT GGT AGC TCC GGA CCG TCG GCC GTC GAC	432
Glu Leu Ala Thr Lys Ala Phe Gly Ser Ser Gly Pro Ser Ala Val Asp	
130 135 140	
AGC GGC ACG GCA ACC GCC CCT CCT GAC CAA TCC TCC GAC GAC GGC GGA	480
Ser Gly Thr Ala Thr Ala Pro Pro Asp Gln Ser Ser Asp Asp Gly Gly	
145 150 155 160	
GCA GGA TCT GAC GTT GAG TCG TAT TCC TCC ATG CCC CCC CTT GAG GGG	528
Ala Gly Ser Asp Val Glu Ser Tyr Ser Ser Met Pro Pro Leu Glu Gly	
165 170 175	
GAG CCG GGG GAC CCC GAT CTC AGC GAC GGG TCT TGG TCT ACC GTG AGT	576
Glu Pro Gly Asp Pro Asp Leu Ser Asp Gly Ser Trp Ser Thr Val Ser	
180 185 190	
GAG GAG GCC GGT GAG GAC GTC GTC TGC TCG ATG TCC TAC ACA TGG	624
Glu Glu Ala Gly Glu Asp Val Val Cys Cys Ser Met Ser Tyr Thr Trp	
195 200 205	

ACA GGC GCT CTG ATC ACG CCA TGC GCT GCG GAG GAA AGC AAG CTG CCC			672
Thr Gly Ala Leu Ile Thr Pro Cys Ala Ala Glu Glu Ser Lys Leu Pro			
210	215	220	
ATC AAC GCG TTG AGC AAC TCT TTG CTG CGT CAC CAC AAC ATG GTC TAC			720
Ile Asn Ala Leu Ser Asn Ser Leu Leu Arg His His Asn Met Val Tyr			
225	230	235	240
GCT ACC ACA TCC CGC AGC GCA AGC CAG CGG CAG AAG AAG GTC ACC TTT			768
Ala Thr Thr Ser Arg Ser Ala Ser Gln Arg Gln Lys Lys Val Thr Phe			
245	250	255	
GAC AGA CTG CAA ATC CTG GAC GAT CAC TAC CAG GAC GTG CTC AAG GAG			816
Asp Arg Leu Gln Ile Leu Asp Asp His Tyr Gln Asp Val Leu Lys Glu			
260	265	270	
ATG AAG GCG AAG GCG TCC ACA GTT AAG GCT AAG CTT CTA TCA GTA GAG			864
Met Lys Ala Lys Ala Ser Thr Val Lys Ala Lys Leu Leu Ser Val Glu			
275	280	285	
GAA GCC TGC AAG CTG ACG CCC CCA CAT TCG GCC AAA TCT AAA TTT GGC			912
Glu Ala Cys Lys Leu Thr Pro Pro His Ser Ala Lys Ser Lys Phe Gly			
290	295	300	
TAT GGG GCA AAG GAC GTC CGG AAC CTA TCC AGC AAG GCC ATT AAC CAC			960
Tyr Gly Ala Lys Asp Val Arg Asn Leu Ser Ser Lys Ala Ile Asn His			
305	310	315	320
ATC CGC TCC GTG TGG GAG GAC TTG TTG GAA GAC ACT GAA ACA CCA ATT			1008
Ile Arg Ser Val Trp Glu Asp Leu Leu Glu Asp Thr Glu Thr Pro Ile			
325	330	335	
GAC ACC ACC ATC ATG GCA AAA AAT GAG GTT TTC TGC GTC CAA CCA GAG			1056
Asp Thr Thr Ile Met Ala Lys Asn Glu Val Phe Cys Val Gln Pro Glu			
340	345	350	

AGA GGA GGC CGC AAG CCA GCT CGC CTT ATC GTG TTC CCA GAC TTG GGG	355	360	365	1104
Arg Gly Gly Arg Lys Pro Ala Arg Leu Ile Val Phe Pro Asp Leu Gly				
GTC CGT GTG TGC GAG AAA ATG GCC CTC TAT GAC GTG GTC TCC ACC CTC	370	375	380	1152
Val Arg Val Cys Glu Lys Met Ala Leu Tyr Asp Val Val Ser Thr Leu				
CCT CAG GCT GTG ATG GGC TCC TCG TAC GGA TTC CAG TAT TCT CCT GGA	385	390	395	1200
Pro Gln Ala Val Met Gly Ser Ser Tyr Gly Phe Gln Tyr Ser Pro Gly				
CAG CGG GTC GAG TTC CTG GTG AAC GCC TGG AAA TCA AAG AAG ACC CCT	405	410	415	1248
Gln Arg Val Glu Phe Leu Val Asn Ala Trp Lys Ser Lys Lys Thr Pro				
ATG GGC TTT GCA TAT GAC ACC CGC TGT TTT GAC TCA ACA GTC ACT GAG	420	425	430	1296
Met Gly Phe Ala Tyr Asp Thr Arg Cys Phe Asp Ser Thr Val Thr Glu				
AAT GAC ATC CGT GTA GAG GAG TCA ATT TAT CAA TGT TGT GAC TTG GCC	435	440	445	1344
Asn Asp Ile Arg Val Glu Glu Ser Ile Tyr Gln Cys Cys Asp Leu Ala				
CCC GAA GCC AGA CAG GCC ATA AGG TCG CTC ACA GAG CGG CTT TAT ATC	450	455	460	1392
Pro Glu Ala Arg Gln Ala Ile Arg Ser Leu Thr Glu Arg Leu Tyr Ile				
GGG GGT CCC CTG ACT AAT TCA AAA GGG CAG AAC TGC GGC TAT CGC CGG	465	470	475	1440
Gly Gly Pro Leu Thr Asn Ser Lys Gly Gln Asn Cys Gly Tyr Arg Arg				
TGC CGC GCG AGC GGC GTG CTG ACG ACT AGC TGC GGT AAT ACC CTC ACA	485	490	495	1488
Cys Arg Ala Ser Gly Val Leu Thr Thr Ser Cys Gly Asn Thr Leu Thr				

TGT TAC TTG AAG GCC TCT GCA GCC TGT CGA GCT GCA AAG CTC CAG GAC 1536  
Cys Tyr Leu Lys Ala Ser Ala Ala Cys Arg Ala Ala Lys Leu Gln Asp  
500 505 510

TGC ACG ATG CTC GTG TGC GGA GAC GGC CTT GTC GTT ATC TGT GAG AGC 1584  
Cys Thr Met Leu Val Cys Gly Asp Asp Leu Val Val Ile Cys Glu Ser  
515 520 525

GCG GGA ACC CAG GAG GAC GCG GCG AGC CTA CGA GTC TTC ACG GAG GCT 1632  
Ala Gly Thr Gln Glu Asp Ala Ala Ser Leu Arg Val Phe Thr Glu Ala  
530 535 540

ATG ACT AGG TAC TCT GCC CCC CCC GGG GAC CCG CCC CAA CCA GAA TAC 1680  
Met Thr Arg Tyr Ser Ala Pro Pro Gly Asp Pro Pro Gln Pro Glu Tyr  
545 550 555 560

GAC CTG GAG TTG ATA ACA TCA TGC TCC TCC AAT GTG TCG GTC GCG CAC 1728  
Asp Leu Glu Leu Ile Thr Ser Cys Ser Ser Asn Val Ser Val Ala His  
565 570 575

GAT GCA TCT GGC AAA AGG GTA TAC TAC CTC ACC CGT GAC CCG 1770  
Asp Ala Ser Gly Lys Arg Val Tyr Tyr Leu Thr Arg Asp Pro  
580 585 590

SEQ ID NO:4

SEQUENCE TYPE:Nucleotide with corresponding protein

SEQUENCE LENGTH:1035 BASE PAIRS

STRANDEDNESS:single

TOPOLOGY:linear

MOLECULE TYPE:cDNA to genomic RNA

ORIGINAL SOURCE ORGANISM:human; serum infectious for  
post-transfusional non-A, non-B hepatitis  
IMMEDIATE EXPERIMENTAL SOURCE:clone JG3 from cDNA library in lambda  
gt11

FEATURES:

from 1 to 1035 bp portion of the PT-NANBH polyprotein

PROPERTIES:probably encodes viral non-structural proteins

ACA GAA GTG GAT GGG GTG CGG CTG CAC AGG TAC GCT CCG GCG TGC AAA 48  
Thr Glu Val Asp Gly Val Arg Leu His Arg Tyr Ala Pro Ala Cys Lys  
5 10 15

CCT CTC CTA CGG GAG GAG GTC ACA TTC CAG GTC GGG CTC AAC CAA TAC 96  
Pro Leu Leu Arg Glu Glu Val Thr Phe Gln Val Gly Leu Asn Gln Tyr  
20 25 30

CTG GTT GGG TCG CAG CTC CCA TGC GAG CCC GAA CCG GAT GTA GCA GTG 144  
Leu Val Gly Ser Gln Leu Pro Cys Glu Pro Glu Pro Asp Val Ala Val  
35 40 45

CTC ACT TCC ATG CTC ACC GAC CCC TCC CAC ATC ACA GCA GAG ACG GCT 192  
Leu Thr Ser Met Leu Thr Asp Pro Ser His Ile Thr Ala Glu Thr Ala  
50 55 60

AAG CGC AGG CTG GCC AGG GGG TCT CCC CCC TCC TTG GCC AGC TCT TCA 240  
Lys Arg Arg Leu Ala Arg Gly Ser Pro Pro Ser Leu Ala Ser Ser Ser  
65 70 75 80

GCT AGC CAG TTG TCT GGC CCT TCC TCG AAG GCG ACA TAC ATT ACC CAA 288  
Ala Ser Gln Leu Ser Gly Pro Ser Ser Lys Ala Thr Tyr Ile Thr Gln  
85 90 95

AAT GAC TTC CCA GAC GCT GAC CTC ATC GAG GCC AAC CTC CTG TGG CGG 336  
Asn Asp Phe Pro Asp Ala Asp Leu Ile Glu Ala Asn Leu Leu Trp Arg  
100 105 110

CAT GAG ATG GGC GGG GAC ATT ACC CGC GTG GAG TCA GAG AAC AAG GTA 384  
His Glu Met Gly Gly Asp Ile Thr Arg Val Glu Ser Glu Asn Lys Val  
115 120 125

GTA ATC CTG GAC TCT TTC GAC CCG CTC CGA GCG GAG GAT GAG CGG 432  
Val Ile Leu Asp Ser Phe Asp Pro Leu Arg Ala Glu Glu Asp Glu Arg  
130 135 140

GAA GTG TCC GTC CCG GCG GAG ATC CTG CGG AAA TCC AAG AAA TTC CCA. 480  
Glu Val Ser Val Pro Ala Glu Ile Leu Arg Lys Ser Lys Lys Phe Pro  
145 150 155 160

CCA GCG ATG CCC GCA TGG GCA CGC CCG GAT TAC AAC CCT CCG CTG CTG 528  
Pro Ala Met Pro Ala Trp Ala Arg Pro Asp Tyr Asn Pro Pro Leu Leu  
165 170 175

GAG TCC TGG AAG GCC CCG GAC TAC GTC CCT CCA GTG GTA CAT GGG TGC 576  
Glu Ser Trp Lys Ala Pro Asp Tyr Val Pro Pro Val Val His Gly Cys  
180 185 190

CCA CTG CCA CCT ACT AAG ACC CCT CCT ATA CCA CCT CCA CGG AGA AAG 624  
Pro Leu Pro Pro Thr Lys Thr Pro Pro Ile Pro Pro Pro Arg Arg Lys  
195 200 205

AGG ACA GTT GTT CTG ACA GAA TCC ACC GTG TCT TCT GCC CTG GCG GAG 672  
Arg Thr Val Val Leu Thr Glu Ser Thr Val Ser Ser Ala Leu Ala Glu  
210 215 220

CTT GCC ACA AAG GCT TTT GGT AGC TCC GGA CCG TCG GCC GTC GAC AGC 720  
Leu Ala Thr Lys Ala Phe Gly Ser Ser Gly Pro Ser Ala Val Asp Ser  
225 230 235 240

GGC ACG GCA ACC GCC CCT CCT GAC CAA TCC TCC GAC GAC GGC GGA GCA 768  
Gly Thr Ala Thr Ala Pro Pro Asp Gln Ser Ser Asp Asp Gly Gly Ala  
245 250 255

GGA TCT GAC GTT GAG TCG TAT TCC TCC ATG CCC CCC CTT GAG GGG GAG 816  
Gly Ser Asp Val Glu Ser Tyr Ser Met Pro Pro Leu Glu Gly Glu  
260 265 270

CCG GGG GAC CCC GAT CTC AGC GAC GGG TCT TGG TCT ACC GTG AGT GAG 864  
Pro Gly Asp Pro Asp Leu Ser Asp Gly Ser Trp Ser Thr Val Ser Glu  
275 280 285

GAG GCC GGT GAG GAC GTC GTC TGC TCG ATG TCC TAC ACA TGG ACA 912  
Glu Ala Gly Glu Asp Val Val Cys Cys Ser Met Ser Tyr Thr Trp Thr  
290 295 300

GGC GCT CTG ATC ACG CCA TGC GCT GCG GAG GAA AGC AAG CTG CCC ATC 960  
Gly Ala Leu Ile Thr Pro Cys Ala Ala Glu Ser Lys Leu Pro Ile  
305 310 315 320

AAC GCG TTG AGC AAC TCT TTG CTG CGT CAC CAC AAC ATG GTC TAC GCT 1008  
Asn Ala Leu Ser Asn Ser Leu Leu Arg His His Asn Met Val Tyr Ala  
325 330 335

ACC ACA TCC CGC AGC GCA AGC CAG CGG 1035  
Thr Thr Ser Arg Ser Ala Ser Gln Arg  
340 345

SEQ ID NO:5

SEQUENCE TYPE:Nucleotide with corresponding protein

SEQUENCE LENGTH: 834 BASE PAIRS

STRANDEDNESS:single

## TOPOLOGY: linear

MOLECULE TYPE:cDNA to genomic RNA

ORIGINAL SOURCE ORGANISM:human; serum infectious for  
post-transfusional non-A, non-B hepatitis  
IMMEDIATE EXPERIMENTAL SOURCE:clone BR11 from cDNA library in lambda  
gt11

## **FEATURES :**

from 1 to 834 bp portion of the PT-NANBH polyprotein

PROPERTIES: probably encodes viral structural proteins

AGA AAA ACC AAA CGT AAC ACC AAC CTC CGC CCA CAG GAC GTC AGG TTC 48  
 Arg Lys Thr Lys Arg Asn Thr Asn Leu Arg Pro Gln Asp Val Arg Phe  
                   5                  10                  15

CCG GGC GGT GGT CAG ATC GTT GGT GGA GTT TAC CTG TTG CCG CGC AGG  
 Pro Gly Gly Gly Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg  
 20 25 30

GGC CCC AGG TTG GGT GTG CGC GCG ACT AGG AAG ACT TCC GAG CGG TCG 144  
 Gly Pro Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser  
 35 40 45

CAA CCT CGT GGA AGG CGA CAA CCT ATC CCC AAG GCT CGC CAG CCC GAG 192  
 Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg Gln Pro Glu  
 50 55 60

GGC AGG GCC TGG GCT CAG CCC GGG TAC CCT TGG CCC CTC TAT GGC AAC 240  
Gly Arg Ala Trp Ala Gln Pro Gly Tyr Pro Trp Pro Leu Tyr Gly Asn  
65 70 75 80

GAG GGC ATG GGG TGG GCA GGA TGG CTC CTG TCA CCC CGT GGC TCC CGG 288  
Glu Gly Met Gly Trp Ala Gly Trp Leu Leu Ser Pro Arg Gly Ser Arg  
85 90 95

CCT AGT TGG GGC CCC ACT GAC CCC CGG CGT AGG TCG CGT AAT TTG GGT 336  
Pro Ser Trp Gly Pro Thr Asp Pro Arg Arg Arg Ser Arg Asn Leu Gly  
100 105 110

AAA GTC ATC GAT ACC CTC ACA TGC GGC TTC GCC GAC TCT CAT GGG GTA 384  
Lys Val Ile Asp Thr Leu Thr Cys Gly Phe Ala Asp Ser His Gly Val  
115 120 125

CAT TCC GCT CGT CGG CGC TCC CTT AGG GGC GCT GCC AGG GCC CTG GCG 432  
His Ser Ala Arg Arg Arg Ser Leu Arg Gly Ala Ala Arg Ala Leu Ala  
130 135 140

CAT GGC GTC CGG GTT CTG GAG GAC GGC GTG AAC TAT GCA ACA GGG AAT 480  
His Gly Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala Thr Gly Asn  
145 150 155 160

TTA CCC GGT TGC TCT TTC TCT ATC TTC CTC TTG GCT TTG CTG TCC TGT 528  
Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu Ser Cys  
165 170 175

TTG ACC ATT CCA GCT TCC GCT TAT GAA GTG CGC AAC GTG TCC GGG ATC 576  
Leu Thr Ile Pro Ala Ser Ala Tyr Glu Val Arg Asn Val Ser Gly Ile  
180 185 190

TAC CAT GTC ACG AAC GAT TGC TCC AAC TCA AGC ATC GTG TAC GAG ACA 624  
Tyr His Val Thr Asn Asp Cys Ser Asn Ser Ser Ile Val Tyr Glu Thr  
195 200 205

GGG GAC ATG ATC ATG CAC ACC CCC GGG TGT GTG CCC TGT GTC CGG GAG	672		
Ala Asp Met Ile Met His Thr Pro Gly Cys Val Pro Cys Val Arg Glu			
210	215	220	
GGT AAT TCC TCC CGC TGC TGG GTA GCG CTC ACT CCC ACG CTC GCG GCC	720		
Gly Asn Ser Ser Arg Cys Trp Val Ala Leu Thr Pro Thr Leu Ala Ala			
225	230	235	240
AAG GAC GCC AGC ATC CCC ACT GCG ACA ATA CGA CGC CAC GTC GAT TTG	768		
Lys Asp Ala Ser Ile Pro Thr Ala Thr Ile Arg Arg His Val Asp Leu			
245	250	255	
CTC GTT GGG GCG GCT GCC TTC TCG TCC GCT ATG TAC GTG GGG GAT CTC	816		
Leu Val Gly Ala Ala Ala Phe Ser Ser Ala Met Tyr Val Gly Asp Leu			
260	265	270	
TGC GGA TCT GTT TTC CCG	834		
Cys Gly Ser Val Phe Pro			
275			

SEQ ID NO:6

SEQUENCE TYPE:Nucleotide

SEQUENCE LENGTH:31 BASES

STRANDEDNESS:single

TOPOLOGY:linear

MOLECULE TYPE:synthetic DNA

ORIGINAL SOURCE ORGANISM:bacteriophage lambda gt11

IMMEDIATE EXPERIMENTAL SOURCE:Oligonucleotide synthesiser; oligo d75

FEATURES:

from 4 to 9 bases BamH1 site

from 10 to 31 bases homologous to upstream portion of lacZ gene  
flanking the EcoR1 site in bacteriophage lambda gt11

from 26 to 31 bases EcoR1 site

PROPERTIES:primes DNA synthesis from the phage vector into cDNA  
inserted at the EcoR1 site and introduces a BamH1 site suitable for  
subsequent cloning into expression vectors.

TAAGGATCCC CCGTCAGTAT CGGCAGAATT C

31

SEQ ID NO:7

SEQUENCE TYPE:Nucleotide

SEQUENCE LENGTH:30 BASES

STRANDEDNESS:single

TOPOLOGY:linear

MOLECULE TYPE:synthetic DNA

ORIGINAL SOURCE ORGANISM:bacteriophage lambda gt11

IMMEDIATE EXPERIMENTAL SOURCE:Oligonucleotide synthesiser; oligo d76

FEATURES:

from 4 to 9 bases BamH1 site

from 10 to 30 bases homologous to downstream portion of lacZ gene  
flanking the EcoR1 site in bacteriophage lambda gt11

PROPERTIES:primes DNA synthesis from the phage vector into cDNA  
inserted at the EcoR1 site and introduces a BamH1 site suitable for  
subsequent cloning into expression vectors.

TATGGATCCG TAGCGACCCGG CGCTCAGCTG

30

SEQ ID NO:8

SEQUENCE TYPE:Nucleotide

SEQUENCE LENGTH:19 BASES

STRANDEDNESS:single

TOPOLOGY:linear

MOLECULE TYPE:synthetic DNA

ORIGINAL SOURCE ORGANISM:human; serum infectious for

post-transfusional non-A, non-B hepatitis

IMMEDIATE EXPERIMENTAL SOURCE:oligonucleotide synthesiser; oligo d94

FEATURES:

from 1 to 19 bases homologous to bases 914 to 932 of the sense strand  
of JG2 (SEQ ID NO : 3)

PROPERTIES:primes DNA synthesis on the negative strand of PT-NANBH  
genomic RNA/DNA.

ATGGGGCAAA GGACGTCCG

19

SEQ ID NO:9

SEQUENCE TYPE:Nucleotide

SEQUENCE LENGTH:24 BASES

STRANDEDNESS:single

TOPOLOGY:linear

MOLECULE TYPE:synthetic DNA

ORIGINAL SOURCE ORGANISM:human; serum infectious for  
post-transfusional non-A, non-B hepatitis  
IMMEDIATE EXPERIMENTAL SOURCE:oligonucleotide synthesiser; oligo d95

FEATURES:

from 1 to 24 bases homologous to bases 1620 to 1643 of the anti-sense  
strand of JG2 (SEQ ID NO : 3)

PROPERTIES:primes DNA synthesis on the positive strand of PT-NANBH  
genomic RNA/DNA.

TACCTAGTCA TAGCCTCCGT GAAG

24

SEQ ID NO:10

SEQUENCE TYPE:Nucleotide

SEQUENCE LENGTH:17 BASES

STRANDEDNESS:single

TOPOLOGY:linear

MOLECULE TYPE:synthetic DNA

ORIGINAL SOURCE ORGANISM:human; serum infectious for  
post-transfusional non-A, non-B hepatitis  
IMMEDIATE EXPERIMENTAL SOURCE:oligonucleotides synthesiser; oligo N1

FEATURES:

from 1 to 17 bases homologous to bases 1033 to 1049 of the sense  
strand of JG2 (SEQ ID NO : 3)

PROPERTIES:primes DNA synthesis on the negative strand of PT-NANBH  
genomic RNA/DNA.

GAGGTTTCT GCGTCCA

17

SEQ ID NO:11

SEQUENCE TYPE:Nucleotide

SEQUENCE LENGTH:17 BASES

STRANDEDNESS:single

TOPOLOGY:linear

MOLECULE TYPE:synthetic DNA

ORIGINAL SOURCE ORGANISM:human; serum infectious for  
post-transfusional non-A, non-B hepatitis  
IMMEDIATE EXPERIMENTAL SOURCE:oligonucleotides synthesiser; oligo N2

FEATURES:

from 1 to 17 bases homologous to bases 1421 to 1437 of the anti-sense  
strand of JG2 (SEQ ID NO : 3)

PROPERTIES:primes DNA synthesis on the positive strand of PT-NANBH  
genomic RNA/DNA.

GCGATAGCCG CAGTTCT

17

SEQ ID NO:12

SEQUENCE TYPE:Nucleotide

SEQUENCE LENGTH:22 BASES

STRANDEDNESS:single

TOPOLOGY:linear

MOLECULE TYPE:synthetic DNA

ORIGINAL SOURCE ORGANISM:human; serum infectious for  
post-transfusional non-A, non-B hepatitis  
IMMEDIATE EXPERIMENTAL SOURCE:oligonucleotide synthesiser; oligo d164

FEATURES:

from 1 to 22 bases homologous to bases 10 to 31 of the sequence in Fig 2 of Okamoto *et al.*, Japan. J. Exp. Med., 1990, 60 167-177, base 22 changed from A to T to introduce Bgl2 recognition site  
from 8 to 13 bases Bgl2 recognition site

PROPERTIES:primes DNA synthesis on the negative strand of PT-NANBH genomic RNA/DNA and introduces a Bgl2 site.

CCACCATAGA TCTCTCCCCT GT

22

SEQ ID NO:13

SEQUENCE TYPE:Nucleotide

SEQUENCE LENGTH:30 BASES

STRANDEDNESS:single

TOPOLOGY:linear

MOLECULE TYPE:synthetic DNA

ORIGINAL SOURCE ORGANISM:human; serum infectious for  
post-transfusional non-A, non-B hepatitis

IMMEDIATE EXPERIMENTAL SOURCE:oligonucleotide synthesiser; oligo d137

FEATURES:

from 1 to 30 bases homologous to bases 154 to 183 of the negative strand of BR11 (SEQ ID NO : 5); bases 174, 177 and 178 modified to introduce an EcoR1 recognition site

from 5 to 10 bases EcoR1 recognition site

PROPERTIES:primes DNA synthesis on the positive strand of PT-NANBH genomic RNA/DNA and introduces an EcoR1 site for cloning

GGGAGAATTC GGGATAGGTT GTCGCCTTCC

30

SEQ ID NO:14

SEQUENCE TYPE:Nucleotide

SEQUENCE LENGTH:27 BASES

STRANDEDNESS:single

TOPOLOGY:linear

MOLECULE TYPE:synthetic DNA

ORIGINAL SOURCE ORGANISM:human; serum infectious for  
post-transfusional non-A, non-B hepatitis  
IMMEDIATE EXPERIMENTAL SOURCE:oligonucleotide synthesiser; oligo d136

FEATURES:

from 1 to 27 bases homologous to bases 672 to 698 of the positive strand of BR11 (SEQ ID NO : 5); base 675 changed to G to introduce an EcoR1 recognition site  
from 4 to 9 bases EcoR1 recognition site

PROPERTIES:primes DNA synthesis on the negative strand of PT-NANBH genomic RNA/DNA and introduces an EcoR1 site for cloning

GGGAATTCC TCCCGCTGCT GGGTAGC

27

SEQ ID NO:15

SEQUENCE TYPE:Nucleotide

SEQUENCE LENGTH:28 BASES

STRANDEDNESS:single

TOPOLOGY:linear

MOLECULE TYPE:synthetic DNA

ORIGINAL SOURCE ORGANISM:chimpanzee; serum infectious for  
post-transfusional non-A, non-B hepatitis  
IMMEDIATE EXPERIMENTAL SOURCE:oligonucleotide synthesiser; oligo d155

FEATURES:

from 1 to 28 bases homologous to bases 462 to 489 of the negative  
strand of figure 47, European Patent Application 88310922.5; bases 483  
and 485 changed to introduce an EcoR1 recognition site  
from 5 to 10 bases EcoR1 recognition site

PROPERTIES:primes DNA synthesis on the positive strand of PT-NANBH  
genomic RNA/DNA and introduces an EcoR1 site for cloning

ACGGGAATTC GACCAGGCAC CTGGGTGT

28

SEQ ID NO:16

SEQUENCE TYPE:Nucleotide

SEQUENCE LENGTH:23 BASES

STRANDEDNESS:single

TOPOLOGY:linear

MOLECULE TYPE:synthetic DNA

ORIGINAL SOURCE ORGANISM:chimpanzee; serum infectious for  
post-transfusional non-A, non-B hepatitis

IMMEDIATE EXPERIMENTAL SOURCE:oligonucleotide synthesiser; oligo d156

FEATURES:

from 1 to 23 bases homologous to bases 3315 to 3337 of the positive  
strand of figure 47, European Patent Application 88310922.5; base 3323  
changed to C to introduce an EcoR1 recognition site  
from 4 to 9 bases EcoR1 recognition site

PROPERTIES:primes DNA synthesis on the negative strand of PT-NANBH  
genomic RNA/DNA and introduces an EcoR1 site for cloning

CTTGAATTCT GGGAGGGCGT CTT

23

SEQ ID NO:17

SEQUENCE TYPE:Nucleotide

SEQUENCE LENGTH:29 BASES

STRANDEDNESS:single

TOPOLOGY:linear

MOLECULE TYPE:synthetic DNA

ORIGINAL SOURCE ORGANISM:human; serum infectious for  
post-transfusional non-A, non-B hepatitis  
IMMEDIATE EXPERIMENTAL SOURCE:oligonucleotide synthesiser; oligo d92

FEATURES:

from 1 to 29 bases homologous to bases 36 to 64 of the negative strand  
of JG2 (SEQ ID NO : 3); bases 57, 58 and 60 changed to introduce an  
EcoR1 recognition site

from 5 to 10 bases EcoR1 recognition site

PROPERTIES:primes DNA synthesis on the positive strand of PT-NANBH  
genomic RNA/DNA and introduces an EcoR1 site for cloning

CGCCGAATTC ATGCCGCCAC AGGAGGTTG

29

SEQ ID NO:18

SEQUENCE TYPE:Nucleotide with corresponding protein

SEQUENCE LENGTH:504 BASE PAIRS

STRANDEDNESS:single

TOPOLOGY:linear

MOLECULE TYPE:cDNA to genomic RNA

ORIGINAL SOURCE ORGANISM:human; serum infectious for  
post-transfusional non-A, non-B hepatitis  
IMMEDIATE EXPERIMENTAL SOURCE:clone 164/137

FEATURES:

from 308 to 504 bp start of the PT-NANBH polyprotein

PROPERTIES:probably encodes viral structural proteins

GATCACTCCC CTGTGAGGAA CTACTGTCTT CACGCAGAAA GCGTCTAGCC ATGGCGTTAG 60  
TATGAGTGTC GTGCAGCCTC CAGGACCCCC CCTCCCGGGA GAGCCATAGT GGTCTGCGGA 120  
ACCGGTGAGT ACACCGGAAT TGCCAGGACG ACCGGGTCCT TTCTTGGATT AACCCGCTCA 180  
ATGCCCTGGAG ATTTGGCGT GCCCCCCCAA GACTGCTAGC CGAGTAGTGT TGGGTGCGGA 240  
AAGGCCTTGT GGTACTGCCT GATAGGGTGC TTGCGAGTGC CCCGGGAGGT CTCGTAGACC 300  
GTGCCACC ATG AGC ACG AAT CCT AAA CCT CAA AGA AAA ACC AAA CGT AAC 349

Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn

5 10

ACC AAC CGC CGC CCA CAG GAC GTC AAG TTC CCG GGC GGT GGT CAG ATC 397  
Thr Asn Pro Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gln Ile  
15 20 25 30

GTT GGT GGA GTT TAC CTG TTG CCG CGC AGG GGC CCC AGG TTG GGT GTG 445  
Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro Arg Leu Gly Val  
35 40 45

CGC GCG ACT AGG AAG ACT TCC GAG CGG TCG CAA CCT CGT GGA AGG CGA 493  
 Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser Gln Pro Arg Gly Arg Arg  
                   50                  55                  60

CAA CCT ATC CC  
Gln Pro Ile Pro  
65

SEQ ID NO:19

SEQUENCE TYPE:Nucleotide with corresponding protein

SEQUENCE LENGTH:1107 BASE PAIRS

## STRANDEDNESS:single

## TOPOLOGY: linear

MOLECULE TYPE:cDNA to genomic RNA

ORIGINAL SOURCE ORGANISM:human; serum infectious for  
post-transfusional non-A, non-B hepatitis  
IMMEDIATE EXPERIMENTAL SOURCE:clone 136/155

## **FEATURES :**

from 1 to 1107 bp portion of the PT-NANBH polyprotein

PROPERTIES: probably encodes viral structural proteins

TCC TCC CGC TGC TGG GTA GCG CTC ACT CCC ACG CTC GCG GCC AAG GAC 48  
 Ser Ser Arg Cys Trp Val Ala Leu Thr Pro Thr Leu Ala Ala Lys Asp  
           5                 10                 15

GCC AGC ATC CCC ACT GCG ACA ATA CGA CGC CAC GTC GAT TTG CTC GTT 96  
 Ala Ser Ile Pro Thr Ala Thr Ile Arg Arg His Val Asp Leu Leu Val  
                   20                  25                  30

GGG GCG GCT GCC TTC TGC TCC GCT ATG TAC GTG GGG GAT CTC TGC GGA 144  
 Gly Ala Ala Ala Phe Cys Ser Ala Met Tyr Val Gly Asp Leu Cys Gly  
                  35                 40                 45

```

TCT GTT TTC CTC GTC TCT CAG CTG TTC ACC TTC TCG CCT CGC CGA CAT      192
Ser Val Phe Leu Val Ser Gln Leu Phe Thr Phe Ser Pro Arg Arg His
      50           55           60

```

CAG ACG GTA CAG GAC TGC AAT TGT TCA ATC TAT CCC GGC CAC GTA TCA 240  
Gln Thr Val Gln Asp Cys Asn Cys Ser Ile Tyr Pro Gly His Val Ser  
65 70 75 80

GGT CAC CGC ATG GCT TGG GAT ATG ATG AAC TGG TCA CCT ACA GCA 288  
Gly His Arg Met Ala Trp Asp Met Met Asn Trp Ser Pro Thr Ala  
85 90 95

GCC CTA GTG GTA TCG CAG CTA CTC CGG ATC CCA CAA GCT GTC GTG GAC 336  
Ala Leu Val Val Ser Gln Leu Leu Arg Ile Pro Gln Ala Val Val Asp  
100 105 110

ATG GTG GCG GGG GCC CAC TGG GGA GTC CTG GCG GGC CTT GCC TAC TAT 384  
Met Val Ala Gly Ala His Trp Gly Val Leu Ala Gly Leu Ala Tyr Tyr  
115 120 125

TCC ATG GTG GGG AAC TGG GCT AAG GTC TTG GTT GTG ATG CTA CTC TTT 432  
Ser Met Val Gly Asn Trp Ala Lys Val Leu Val Val Met Leu Leu Phe  
130 135 140

GCC GGC GTT GAC GGG GAA CCT TAC ACG ACA GGG GGG ACA CAC GGC CGC 480  
Ala Gly Val Asp Gly Glu Pro Tyr Thr Thr Gly Gly Thr His Gly Arg  
145 150 155 160

GCC GCC CAC GGG CTT ACA TCC CTC TTC ACA CCT GGG CCG GCT CAG AAA 528  
Ala Ala His Gly Leu Thr Ser Leu Phe Thr Pro Gly Pro Ala Gln Lys  
165 170 175

ATC CAG CTT GTA AAC ACC AAC GGC AGC TGG CAC ATC AAC AGA ACT GCC 576  
Ile Gln Leu Val Asn Thr Asn Gly Ser Trp His Ile Asn Arg Thr Ala  
180 185 190

TTG AAC TGC AAT GAC TCC CTC CAA ACT GGG TTC CTT GCC GCG CTG TTC 624  
Leu Asn Cys Asn Asp Ser Leu Gln Thr Gly Phe Leu Ala Ala Leu Phe  
195 200 205

TAC ACG CAC AGG TTC AAT GCG TCC GGA TGC TCA GAG CGC ATG GCC AGC 672  
Tyr Thr His Arg Phe Asn Ala Ser Gly Cys Ser Glu Arg Met Ala Ser  
210 215 220

TGC CGC CCC ATT GAC CAG TTC GAT CAG GGG TGG GGT CCC ATC ACT TAT 720  
Cys Arg Pro Ile Asp Gln Phe Asp Gln Gly Trp Gly Pro Ile Thr Tyr  
225 230 235 240

AAT GAG TCC CAC GGC TTG GAC CAG AGG CCC TAT TGC TGG CAC TAC GCA 768  
Asn Glu Ser His Gly Leu Asp Gln Arg Pro Tyr Cys Trp His Tyr Ala  
245 250 255

CCT CAA CCG TGT GGT ATC GTG CCC GCG TTG CAG GTG TGT GGC CCA GTG 816  
Pro Gln Pro Cys Gly Ile Val Pro Ala Leu Gln Val Cys Gly Pro Val  
260 265 270

TAC TGT TTC ACT CCA AGC CCT GTT GTG GTG GGG ACG ACC GAT CGT TTC 864  
Tyr Cys Phe Thr Pro Ser Pro Val Val Val Gly Thr Thr Asp Arg Phe  
275 280 285

GGC GCC CCT ACG TAC AGA TGG GGT GAG AAT GAG ACG GAC GTG CTG CTT 912  
Gly Ala Pro Thr Tyr Arg Trp Gly Glu Asn Glu Thr Asp Val Leu Leu  
290 295 300

CTC AAC AAC ACG CGG CCG CCA CGG GGC AAC TGG TTC GGC TGT ACA TGG 960  
Leu Asn Asn Thr Arg Pro Pro Arg Gly Asn Trp Phe Gly Cys Thr Trp  
305 310 315 320

ATG AAT AGC ACC GGG TTC ACC AAG ACG TGT GGG GGC CCC CCG TGC AAC 1008  
Met Asn Ser Thr Gly Phe Thr Lys Thr Cys Gly Gly Pro Pro Cys Asn  
325 330 335

ATC GGG GGG GTC GGC AAC AAC ACT TTG ATC TGC CCC ACG GAC TGC TTC 1056  
Ile Gly Gly Val Gly Asn Asn Thr Leu Ile Cys Pro Thr Asp Cys Phe  
340 345 350

CGG AAG CAT CCC GAG GCC ACT TAC ACC AAA TGC GGT TCG GGG CCT TGG 1104  
Arg Lys His Pro Glu Ala Thr Tyr Thr Lys Cys Gly Ser Gly Pro Trp  
355 360 365

TTG

1107

Leu

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.26.253600

SEQ ID NO:20

SEQUENCE TYPE:Nucleotide with corresponding protein

SEQUENCE LENGTH: 2043 BASE PAIRS

STRANDEDNESS:single

## TOPOLOGY: linear

MOLECULE TYPE:cDNA to genomic RNA

ORIGINAL SOURCE ORGANISM:human; serum infectious for  
post-transfusional non-A, non-B hepatitis  
IMMEDIATE EXPERIMENTAL SOURCE:clone 156/92

## **FEATURES :**

from 1 to 2043 bp portion of the PT-NANBH polyprotein

**PROPERTIES:** probably encodes viral non-structural proteins

TGG GAG GGC GTC TTC ACA GCC CTC ACC CAC GTG GAT GCC CAC TTC CTG 48  
Trp Glu Gly Val Phe Thr Gly Leu Thr His Val Asp Ala His Phe Leu  
5 10 15

TCC CAA ACA AAG CAG GCA GGA GAC AAC TTC CCC TAC CTG GTG GCG TAC 96  
 Ser Gln Thr Lys Gln Ala Gly Asp Asn Phe Pro Tyr Leu Val Ala Tyr  
           20                  25                  30

CAG GCT ACT GTG TGC GCT AGG GCC CAG GCC CCA CCT CCA TCA TGG GAT 144  
 Gln Ala Thr Val Cys Ala Arg Ala Gln Ala Pro Pro Pro Ser Trp Asp  
                  35                40                45

CAA ATG TCG AAG TGT CTC ATA CGG CTA AAG CCT ACT CTG CGC GGG CCA 192  
 Gln Met Trp Lys Cys Leu Ile Arg Leu Lys Pro Thr Leu Arg Gly Pro  
 50 55 60

ACA CCC TTG CTG TAT AGG CTG GGA GCC GTC CAA AAC GAG GTC ACC CTC	240
Thr Pro Leu Leu Tyr Arg Leu Gly Ala Val Gln Asn Glu Val Thr Leu	
65 70 75 80	
ACA CAC CCC ATA ACC AAA TTC ATC ATG GCA TGC ATG TCA GCC GAC CTG	288
Thr His Pro Ile Thr Lys Phe Ile Met Ala Cys Met Ser Ala Asp Leu	
85 90 95	
GAG GTC GTC ACG AGC ACC TGG GTG CTG GTG GGC GGG GTC CTT GCA GCT	336
Glu Val Val Thr Ser Thr Trp Val Leu Val Gly Gly Val Leu Ala Ala	
100 105 110	
CTG GCT GCG TAT TGC TTG ACA ACA GGC AGC GTG GTC ATT GTG GGT AGG	384
Leu Ala Ala Tyr Cys Leu Thr Thr Gly Ser Val Val Ile Val Gly Arg	
115 120 125	
ATC ATC TTG TCC GGG CGG CCG GCT ATT GTT CCC GAC AGG GAA GTC CTC	432
Ile Ile Leu Ser Gly Arg Pro Ala Ile Val Pro Asp Arg Glu Val Leu	
130 135 140	
TAC CAG GAG TTC GAT GAG ATG GAA GAG TGC GCG TCG CAC CTC CCT TAC	480
Tyr Gln Glu Phe Asp Glu Met Glu Glu Cys Ala Ser His Leu Pro Tyr	
145 150 155 160	
ATC GAG CAG GGA ATG CAG CTC GCC GAG CAG TTC AAG CAA AAA GCG CTC	528
Ile Glu Gln Gly Met Gln Leu Ala Glu Gln Phe Lys Gln Lys Ala Leu	
165 170 175	
GGG TTG CTG CAG ACA GCC ACC AAG CAA GCG GAG GCC GCT GCT CCC GTG	576
Gly Leu Leu Gln Thr Ala Thr Lys Gln Ala Glu Ala Ala Ala Pro Val	
180 185 190	
GTG GAG TCC AAG TGG CGA GCC CTT GAG ACC TTC TGG GCG AAA CAC ATG	624
Val Glu Ser Lys Trp Arg Ala Leu Glu Thr Phe Trp Ala Lys His Met	
195 200 205	

TGG AAC TTC ATC AGC GGG ATA CAG TAC TTA GCA GGC TTG TCC ACT CTG 672  
Trp Asn Phe Ile Ser Gly Ile Gln Tyr Leu Ala Gly Leu Ser Thr Leu  
210 215 220

CCT GGG AAT CCC GCG ATT GCA TCA CTG ATG GCG TTC ACA GCC TCT GTC 720  
Pro Gly Asn Pro Ala Ile Ala Ser Leu Met Ala Phe Thr Ala Ser Val  
225 230 235 240

ACT AGC CCG CTC ACC ACC CAA TCT ACC CTC CTG CTT AAC ATC CTG GGG 768  
Thr Ser Pro Leu Thr Thr Gln Ser Thr Leu Leu Leu Asn Ile Leu Gly  
245 250 255

GGA TGG GTA GCC GCC CAA CTC GCT CCC CCC AGT GCT GCT TCA GCT TTC 816  
Gly Trp Val Ala Ala Gln Leu Ala Pro Pro Ser Ala Ala Ser Ala Phe  
260 265 270

GTA GGC GCC GGC ATT GCT GGT GCG GCT GTT GGC AGC ATA GGC CTT GGG 864  
Val Gly Ala Gly Ile Ala Gly Ala Ala Val Gly Ser Ile Gly Leu Gly  
275 280 285

AAG GTG CTT GTG GAC ATC TTG GCG GGC TAT GGA GCA GGA GTG GCA GGC 912  
Lys Val Leu Val Asp Ile Leu Ala Gly Tyr Gly Ala Gly Val Ala Gly  
290 295 300

GCG CTC GTG GCC TTT AAG GTC ATG AGC GGC GAA ATG CCC TCC ACC GAG 960  
Ala Leu Val Ala Phe Lys Val Met Ser Gly Glu Met Pro Ser Thr Glu  
305 310 315 320

GAC CTG GTT AAC TTA CTC CCT GCC ATC CTC TCT CCT GGT GCC CTG GTC 1008  
Asp Leu Val Asn Leu Leu Pro Ala Ile Leu Ser Pro Gly Ala Leu Val  
325 330 335

GTC GGG GTC GTG TGC GCA GCG ATA CTG CGT CGG CAC GTG GGT CCA GGG 1056  
Val Gly Val Val Cys Ala Ala Ile Leu Arg Arg His Val Gly Pro Gly  
340 345 350

GAG GGG GCT GTG CAG TGG ATG AAC CGG CTG ATA GCG TTC GCC TCG CGG 1104  
Glu Gly Ala Val Gln Trp Met Asn Arg Leu Ile Ala Phe Ala Ser Arg  
355 360 365

GGT AAC CAT GTT TCC CCC ACG CAC TAT GTG CCA GAG AGC GAC GCC GCA 1152  
Gly Asn His Val Ser Pro Thr His Tyr Val Pro Glu Ser Asp Ala Ala  
370 375 380

GCA CGT GTC ACT CAG ATC CTC TCC GAC CTT ACT ATC ACC CAA CTG TTG 1200  
Ala Arg Val Thr Gln Ile Leu Ser Asp Leu Thr Ile Thr Gln Leu Leu  
385 390 395 400

AAG AGG CTC CAC CAG TGG ATT AAC GAG GAC TGC TCC ACG CCC TGC TCC 1248  
Lys Arg Leu His Gln Trp Ile Asn Glu Asp Cys Ser Thr Pro Cys Ser  
405 410 415

GGC TCG TGG CTA AGG GAT GTT TGG GAC TGG ATA TGC ACA GTT TTG GCT 1296  
Gly Ser Trp Leu Arg Asp Val Trp Asp Trp Ile Cys Thr Val Leu Ala  
420 425 430

GAC TTC AAG ACC TGG CTC CAG TCC AAG CTC CTG CCG CGA TTA CCG GGA 1344  
Asp Phe Lys Thr Trp Leu Gln Ser Lys Leu Leu Pro Arg Leu Pro Gly  
435 440 445

GTC CCC TTT TTC TCA TGC CAA CGT GGG TAC AAG GGG GTC TGG CGG GGA 1392  
Val Pro Phe Phe Ser Cys Gln Arg Gly Tyr Lys Gly Val Trp Arg Gly  
450 455 460

GAC GGC ATC ATG CAG ACC ACC TGC TCA TGT GGA GCA CAG ATC ACC GGA 1440  
Asp Gly Ile Met Gln Thr Thr Cys Ser Cys Gly Ala Gln Ile Thr Gly  
465 470 475 480

CAT GTC AAA AAC GGT TCC ATG AGG ATC GTT GGG CCT AAG ACC TGT AGT 1488  
His Val Lys Asn Gly Ser Met Arg Ile Val Gly Pro Lys Thr Cys Ser  
485 490 495

AAC ATG TGG CAT GGA ACA TTC CCC ATC AAC GCA TAC ACC ACG GGC CCC 1536  
Asn Met Trp His Gly Thr Phe Pro Ile Asn Ala Tyr Thr Thr Gly Pro  
500 505 510

TGC ACG CCC TCC CCA GCG CCA AAC TAT TCC AGG GCG CTG TGG CGG GTG 1584  
Cys Thr Pro Ser Pro Ala Pro Asn Tyr Ser Arg Ala Leu Trp Arg Val  
515 520 525

GCT GCT GAG GAG TAC GTG GAG GTT ACG CGG GTG GGG GAT TTC CAC TAC 1632  
Ala Ala Glu Glu Tyr Val Glu Val Thr Arg Val Gly Asp Phe His Tyr  
530 535 540

GTG ACG AGC ATG ACC ACT GAC AAC GTA AAA TGC CCG TGC CAG GTT CCA 1680  
Val Thr Ser Met Thr Thr Asp Asn Val Lys Cys Pro Cys Gln Val Pro  
545 550 555 560

GCC CCC GAA TTC TTC ACA GAA GTG GAT GGG GTG CGG CTG CAC AGG TAC 1728  
Ala Pro Glu Phe Phe Thr Glu Val Asp Gly Val Arg Leu His Arg Tyr  
565 570 575

GCT CCG GCG TGC AAA CCT CTC CTA CGG GAG GTC ACA TTC CAG GTC 1776  
Ala Pro Ala Cys Lys Pro Leu Leu Arg Glu Glu Val Thr Phe Gln Val  
580 585 590

GGG CTC AAC CAA TAC CTG GTT GGG TCG CAG CTC CCA TGC GAG CCC GAA 1824  
Gly Leu Asn Gln Tyr Leu Val Gly Ser Gln Leu Pro Cys Glu Pro Glu  
595 600 605

CCG GAT GTA GCA GTG CTC ACT TCC ATG CTC ACC GAC CCC TCC CAC ATC 1872  
Pro Asp Val Ala Val Leu Thr Ser Met Leu Thr Asp Pro Ser His Ile  
610 615 620

ACA GCA GAG ACG GCT AAG CGC AGG CTG GCC AGG GGG TCT CCC CCC TCC 1920  
Thr Ala Glu Thr Ala Lys Arg Arg Leu Ala Arg Gly Ser Pro Pro Ser  
625 630 635 640

TTG GCC AGC TCT TCA GCT AGC CAG TTG TCT GCG CCT TCC TCG AAG GCG 1968  
Leu Ala Ser Ser Ser Ala Ser Gln Leu Ser Ala Pro Ser Ser Lys Ala  
645 650 655

ACA TAC ATT ACC CAA AAT GAC TTC CCA GAC GCT GAC CTC ATC GAG GCC 2016  
Thr Tyr Ile Thr Gln Asn Asp Phe Pro Asp Ala Asp Leu Ile Glu Ala  
660 665 670

AAC CTC CTG TGG CGG CAT GAG ATG GGC 2043  
Asn Leu Leu Trp Arg His Glu Met Gly  
675 680

© 2024 OpenStax. All Rights Reserved.

SEQ ID NO:21

SEQUENCE TYPE:Nucleotide with corresponding protein

SEQUENCE LENGTH:2116 BASE PAIRS

STRANDEDNESS:single

TOPOLOGY:linear

MOLECULE TYPE:cDNA to genomic RNA

ORIGINAL SOURCE ORGANISM:human; serum infectious for  
post-transfusional non-A, non-B hepatitis

IMMEDIATE EXPERIMENTAL SOURCE:contig formed by cDNA clones from 5' end  
of the genome

FEATURES:

from 308 to 2116 bp start of the PT-NANBH polyprotein

PROPERTIES:viral structural and non-structural proteins

GATCACTCCC CTGTGAGGAA CTACTGTCTT CACGCAGAAA GCGTCTAGCC ATGGCGTTAG 60  
TATGAGTGTC GTGCAGCCTC CAGGACCCCC CCTCCCGGGA GAGCCATAAGT GGTCTGCGGA 120  
ACCGGTGAGT ACACCGGAAT TGCCAGGACG ACCGGGTCCCT TTCTTGGATT AACCCGCTCA 180  
ATGCCTGGAG ATTTGGCGT GCCCCCGCAA GACTGCTAGC CGACTAGTGT TGGGTGCGGA 240  
AAGGCCTTGT GGTACTGCCCT GATAAGGTGC TTGCGAGTGC CCCGGGAGGT CTCGTAGACC 300  
GTGCCACC ATG AGC ACG AAT CCT AAA CCT CAA AGA AAA ACC AAA CGT AAC 349

Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn

5 10

ACC AAC CGC CGC CCA CAG GAC GTC AAG TTC CCG GGC GGT GGT CAG ATC 397  
Thr Asn Pro Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gln Ile  
15 20 25 30

GTT GGT GGA GTT TAC CTG TTG CCG CCC AGG GGC CCC AGG TTG GGT GTG 445  
Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro Arg Leu Gly Val  
35 40 45

CGC GCG ACT AGG AAG ACT TCC GAG CGG TCG CAA CCT CGT GGA AGG CGA	493	
Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser Gln Pro Arg Gly Arg Arg		
50	55	60
CAA CCT ATC CCC AAG GCT CGC CAG CCC GAG GGC AGG GCC TGG GCT CAG	541	
Gln Pro Ile Pro Lys Ala Arg Gln Pro Glu Gly Arg Ala Trp Ala Gln		
65	70	75
CCC GGG TAC CCT TGG CCC CTC TAT GGC AAC GAG GGC ATG GGG TGG GCA	589	
Pro Gly Tyr Pro Trp Pro Leu Tyr Gly Asn Glu Gly Met Gly Trp Ala		
80	85	90
GGA TGG CTC CTG TCA CCC CGT GGC TCC CGG CCT AGT TGG GGC CCC ACT	637	
Gly Trp Leu Leu Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Thr		
100	105	110
115		
GAC CCC CGG CGT AGG TCG CGT AAT TTG GGT AAA GTC ATC GAT ACC CTC	685	
Asp Pro Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu		
120	125	130
ACA TGC GGC TTC GCC GAC CTC ATG GGG TAC ATT CCG CTC GTC GGC GCT	733	
Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val Gly Ala		
135	140	145
CCC TTA GGG GGC GCT GCC AGG GCC CTG GCG CAT GGC GTC CGG GTT CTG	781	
Pro Leu Gly Gly Ala Ala Arg Ala Leu Ala His Gly Val Arg Val Leu		
150	155	160
GAG GAC GGC GTG AAC TAT GCA ACA CGG AAT TTA CCC GGT TGC TCT TTC	829	
Glu Asp Gly Val Asn Tyr Ala Thr Gly Asn Leu Pro Gly Cys Ser Phe		
165	170	175
TCT ATC TTC CTC TTG GCT TTG CTG TCC TGT TTG ACC ATT CCA GCT TCC	877	
Ser Ile Phe Leu Leu Ala Leu Leu Ser Cys Leu Thr Ile Pro Ala Ser		
180	185	190
195		

GCT TAT GAA GTG CGC AAC GTG TCC GGG ATC TAC CAT GTC ACG AAC GAT 925  
Ala Tyr Glu Val Arg Asn Val Ser Gly Ile Tyr His Val Thr Asn Asp  
200 205 210

TGC TCC AAC TCA AGC ATC GTG TAC GAG ACA GCG GAC ATG ATC ATG CAC 973  
Cys Ser Asn Ser Ser Ile Val Tyr Glu Thr Ala Asp Met Ile Met His  
215 220 225

ACC CCC GGG TGT GTG CCC TGT GTC CGG GAG GGT AAT TCC TCC CGC TGC 1021  
Thr Pro Gly Cys Val Pro Cys Val Arg Glu Gly Asn Ser Ser Arg Cys  
230 235 240

TGG GTA GCG CTC ACT CCC ACG CTC GCG GCC AAG GAC GCC AGC ATC CCC 1069  
Trp Val Ala Leu Thr Pro Thr Leu Ala Ala Lys Asp Ala Ser Ile Pro  
245 250 255

ACT GCG ACA ATA CGA CGC CAC GTC GAT TTG CTC GTT GGG GCG GCT GCC 1117  
Thr Ala Thr Ile Arg Arg His Val Asp Leu Leu Val Gly Ala Ala Ala  
260 265 270 275

TTC TGC TCC GCT ATG TAC GTG GGG GAT CTC TGC GGA TCT GTT TTC CTC 1165  
Phe Cys Ser Ala Met Tyr Val Gly Asp Leu Cys Gly Ser Val Phe Leu  
280 285 290

GTC TCT CAG CTG TTC ACC TTC TCG CCT CGC CGA CAT CAG ACG GTA CAG 1213  
Val Ser Gln Leu Phe Thr Phe Ser Pro Arg Arg His Gln Thr Val Gln  
295 300 305

GAC TGC AAT TGT TCA ATC TAT CCC GGC CAC GTA TCA GGT CAC CGC ATG 1261  
Asp Cys Asn Cys Ser Ile Tyr Pro Gly His Val Ser Gly His Arg Met  
310 315 320

GCT TGG GAT ATG ATG ATG AAC TGG TCA CCT ACA GCA GCC CTA GTG GTA 1309  
Ala Trp Asp Met Met Asn Trp Ser Pro Thr Ala Ala Leu Val Val  
325 330 335

TCG CAG CTA CTC CGG ATC CCA CAA GCT GTC GTG GAC ATG GTG GCG GGG 1357  
Ser Gln Leu Leu Arg Ile Pro Gln Ala Val Val Asp Met Val Ala Gly  
340 345 350 355

GCC CAC TGG GGA GTC CTG GCG GGC CTT GCC TAC TAT TCC ATG GTG GGG 1405  
Ala His Trp Gly Val Leu Ala Gly Leu Ala Tyr Tyr Ser Met Val Gly  
360 365 370

AAC TGG GCT AAG GTC TTG GTT GTG ATG CTA CTC TTT GCC GGC GTT GAC 1453  
Asn Trp Ala Lys Val Leu Val Val Met Leu Leu Phe Ala Gly Val Asp  
375 380 385

GGG GAA CCT TAC ACG ACA GGG GGG ACA CAC GGC CGC GCC GCC CAC GGG 1501  
Gly Glu Pro Tyr Thr Thr Gly Gly Thr His Gly Arg Ala Ala His Gly  
390 395 400

CTT ACA TCC CTC TTC ACA CCT GGG CCG GCT CAG AAA ATC CAG CTT GTA 1549  
Leu Thr Ser Leu Phe Thr Pro Gly Pro Ala Gln Lys Ile Gln Leu Val  
405 410 415

AAC ACC AAC GGC AGC TGG CAC ATC AAC AGA ACT GCC TTG AAC TGC AAT 1597  
Asn Thr Asn Gly Ser Trp His Ile Asn Arg Thr Ala Leu Asn Cys Asn  
420 425 430 435

GAC TCC CTC CAA ACT GGG TTC CTT GCC GCG CTG TTC TAC ACG CAC AGG 1645  
Asp Ser Leu Gln Thr Gly Phe Leu Ala Ala Leu Phe Tyr Thr His Arg  
440 445 450

TTC AAT GCG TCC GGA TGC TCA GAG CGC ATG GCC AGC TGC CGC CCC ATT 1693  
Phe Asn Ala Ser Gly Cys Ser Glu Arg Met Ala Ser Cys Arg Pro Ile  
455 460 465

GAC CAG TTC GAT CAG GGG TGG GGT CCC ATC ACT TAT AAT GAG TCC CAC 1741  
Asp Gln Phe Asp Gln Gly Trp Gly Pro Ile Thr Tyr Asn Glu Ser His  
470 475 480

GGC TTG GAC CAG AGG CCC TAT TGC TGG CAC TAC GCA CCT CAA CCG TGT 1789  
Gly Leu Asp Gln Arg Pro Tyr Cys Trp His Tyr Ala Pro Gln Pro Cys  
485 490 495

GGT ATC GTG CCC GCG TTG CAG GTG TGT GGC CCA GTG TAC TGT TTC ACT 1837  
Gly Ile Val Pro Ala Leu Gln Val Cys Gly Pro Val Tyr Cys Phe Thr  
500 505 510 515

CCA AGC CCT GTT GTG GTG GGG ACG ACC GAT CGT TTC GGC GCC CCT ACG 1885  
Pro Ser Pro Val Val Val Gly Thr Thr Asp Arg Phe Gly Ala Pro Thr  
520 525 530

TAC AGA TGG GGT GAG AAT GAG ACG GAC GTG CTG CTT CTC AAC AAC ACG 1933  
Tyr Arg Trp Gly Glu Asn Glu Thr Asp Val Leu Leu Asn Asn Thr  
535 540 545

CGG CCG CCA CGG GGC AAC TGG TTC GGC TGT ACA TGG ATG AAT AGC ACC 1981  
Arg Pro Pro Arg Gly Asn Trp Phe Gly Cys Thr Trp Met Asn Ser Thr  
550 555 560

GGG TTC ACC AAG ACG TGT GGG GGC CCC CCG TGC AAC ATC GGG GGG GTC 2029  
Gly Phe Thr Lys Thr Cys Gly Gly Pro Pro Cys Asn Ile Gly Gly Val  
565 570 575

GGC AAC AAC ACT TTG ATC TGC CCC ACG GAC TGC TTC CGG AAG CAT CCC 2077  
Gly Asn Asn Thr Leu Ile Cys Pro Thr Asp Cys Phe Arg Lys His Pro  
580 585 590 595

GAG GCC ACT TAC ACC AAA TGC GGT TCG GGG CCT TGG TTG 2116  
Glu Ala Thr Tyr Thr Lys Cys Gly Ser Gly Pro Trp Leu  
600 605

SEQ ID NO:22

SEQUENCE TYPE:Nucleotide with corresponding protein

SEQUENCE LENGTH:3750 BASE PAIRS

STRANDEDNESS:single

TOPOLOGY:linear

MOLECULE TYPE:cDNA to genomic RNA

ORIGINAL SOURCE ORGANISM:human; serum infectious for  
post-transfusional non-A, non-B hepatitis  
IMMEDIATE EXPERIMENTAL SOURCE:contig formed by cDNA clones from 3' end  
of the genome

FEATURES:

from 1 to 3750 bp portion of the PT-NANBH polyprotein

PROPERTIES:viral non-structural proteins

TGG GAG GGC GTC TTC ACA GGC CTC ACC CAC GTG GAT GCC CAC TTC CTG	48	
Trp Glu Gly Val Phe Thr Gly Leu Thr His Val Asp Ala His Phe Leu		
5	10	15
Ser Gln Thr Lys Gln Ala Gly Asp Asn Phe Pro Tyr Leu Val Ala Tyr	96	
20	25	30
CAG GCT ACT GTG TGC GCT AGG GCC CAG GCC CCA CCT CCA TCA TGG GAT	144	
Gln Ala Thr Val Cys Ala Arg Ala Gln Ala Pro Pro Pro Ser Trp Asp		
35	40	45
CAA ATG TGG AAG TGT CTC ATA CGG CTA AAG CCT ACT CTG CGC GGG CCA	192	
Gln Met Trp Lys Cys Leu Ile Arg Leu Lys Pro Thr Leu Arg Gly Pro		
50	55	60

ACA CCC TTG CTG TAT AGG CTG GGA GCC GTC CAA AAC GAG GTC ACC CTC 240  
Thr Pro Leu Leu Tyr Arg Leu Gly Ala Val Gln Asn Glu Val Thr Leu  
65 70 75 80

ACA CAC CCC ATA ACC AAA TTC ATC ATG GCA TGC ATG TCA GCC GAC CTG 288  
Thr His Pro Ile Thr Lys Phe Ile Met Ala Cys Met Ser Ala Asp Leu  
85 90 95

GAG GTC GTC ACG AGC ACC TGG GTG CTG GTG GGC GGG GTC CTT GCA GCT 336  
Glu Val Val Thr Ser Thr Trp Val Leu Val Gly Gly Val Leu Ala Ala  
100 105 110

CTG GCT GCG TAT TGC TTG ACA ACA GGC AGC GTG GTC ATT GTG GGT AGG 384  
Leu Ala Ala Tyr Cys Leu Thr Thr Gly Ser Val Val Ile Val Gly Arg  
115 120 125

ATC ATC TTG TCC GGG CGG CCG GCT ATT GTT CCC GAC AGG GAA GTC CTC 432  
Ile Ile Leu Ser Gly Arg Pro Ala Ile Val Pro Asp Arg Glu Val Leu  
130 135 140

TAC CAG GAG TTC GAT GAG ATG GAA GAG TGC GCG TCG CAC CTC CCT TAC 480  
Tyr Gln Glu Phe Asp Glu Met Glu Glu Cys Ala Ser His Leu Pro Tyr  
145 150 155 160

ATC GAG CAG GGA ATG CAG CTC GCC GAG CAG TTC AAG CAA AAA GCG CTC 528  
Ile Glu Gln Gly Met Gln Leu Ala Glu Gln Phe Lys Gln Lys Ala Leu  
165 170 175

GGG TTG CTG CAG ACA GCC ACC AAG CAA GCG GAG GCC GCT GCT CCC GTG 576  
Gly Leu Leu Gln Thr Ala Thr Lys Gln Ala Glu Ala Ala Pro Val  
180 185 190

GTG GAG TCC AAG TGG CGA GCC CTT GAG ACC TTC TGG GCG AAA CAC ATG 624  
Val Glu Ser Lys Trp Arg Ala Leu Glu Thr Phe Trp Ala Lys His Met  
195 200 205

TGG AAC TTC ATC AGC GGG ATA CAG TAC TTA GCA GGC TTG TCC ACT CTG	672
Trp Asn Phe Ile Ser Gly Ile Gln Tyr Leu Ala Gly Leu Ser Thr Leu	
210 215 220	
CCT GGG AAT CCC GCG ATT GCA TCA CTG ATG GCG TTC ACA GCC TCT GTC	720
Pro Gly Asn Pro Ala Ile Ala Ser Leu Met Ala Phe Thr Ala Ser Val	
225 230 235 240	
ACT AGC CCG CTC ACC ACC CAA TCT ACC CTC CTG CTT AAC ATC CTG GGG	768
Thr Ser Pro Leu Thr Thr Gln Ser Thr Leu Leu Leu Asn Ile Leu Gly	
245 250 255	
GGA TGG GTA GCC GCC CAA CTC GCT CCC CCC AGT GCT GCT TCA GCT TTC	816
Gly Trp Val Ala Ala Gln Leu Ala Pro Pro Ser Ala Ala Ser Ala Phe	
260 265 270	
GTA GGC GCC GGC ATT GCT GGT GCG GCT GTT GGC AGC ATA GGC CTT GGG	864
Val Gly Ala Gly Ile Ala Gly Ala Ala Val Gly Ser Ile Gly Leu Gly	
275 280 285	
AAG GTG CTT GTG GAC ATC TTG GCG GGC TAT GGA GCA GGA GTG GCA GGC	912
Lys Val Leu Val Asp Ile Leu Ala Gly Tyr Gly Ala Gly Val Ala Gly	
290 295 300	
GCG CTC GTG GCC TTT AAG GTC ATG AGC GGC GAA ATG CCC TCC ACC GAG	960
Ala Leu Val Ala Phe Lys Val Met Ser Gly Glu Met Pro Ser Thr Glu	
305 310 315 320	
GAC CTG GTT AAC TTA CTC CCT GCC ATC CTC TCT CCT GGT GCC CTG GTC	1008
Asp Leu Val Asn Leu Leu Pro Ala Ile Leu Ser Pro Gly Ala Leu Val	
325 330 335	
GTC GGG GTC GTG TGC GCA GCG ATA CTG CGT CGG CAC GTG GGT CCA GGG	1056
Val Gly Val Val Cys Ala Ala Ile Leu Arg Arg His Val Gly Pro Gly	
340 345 350	

GAG GGG GCT GTG CAG TGG ATG AAC CGG CTG ATA GCG TTC GCC TCG CGG 1104  
Glu Gly Ala Val Gln Trp Met Asn Arg Leu Ile Ala Phe Ala Ser Arg  
355 360 365

GGT AAC CAT GTT TCC CCC ACG CAC TAT GTG CCA GAG AGC GAC GCC GCA 1152  
Gly Asn His Val Ser Pro Thr His Tyr Val Pro Glu Ser Asp Ala Ala  
370 375 380

GCA CGT GTC ACT CAG ATC CTC TCC GAC CTT ACT ATC ACC CAA CTG TTG 1200  
Ala Arg Val Thr Gln Ile Leu Ser Asp Leu Thr Ile Thr Gln Leu Leu  
385 390 395 400

AAG AGG CTC CAC CAG TGG ATT AAC GAG GAC TGC TCC ACG CCC TGC TCC 1248  
Lys Arg Leu His Gln Trp Ile Asn Glu Asp Cys Ser Thr Pro Cys Ser  
405 410 415

GGC TCG TGG CTA AGG GAT GTT TGG GAC TGG ATA TGC ACA GTT TTG GCT 1296  
Gly Ser Trp Leu Arg Asp Val Trp Asp Trp Ile Cys Thr Val Leu Ala  
420 425 430

GAC TTC AAG ACC TGG CTC CAG TCC AAG CTC CTG CCG CGA TTA CCG GGA 1344  
Asp Phe Lys Thr Trp Leu Gln Ser Lys Leu Leu Pro Arg Leu Pro Gly  
435 440 445

GTC CCC TTT TTC TCA TGC CAA CGT GGG TAC AAG GGG GTC TGG CGG GGA 1392  
Val Pro Phe Phe Ser Cys Gln Arg Gly Tyr Lys Gly Val Trp Arg Gly  
450 455 460

GAC GGC ATC ATG CAG ACC ACC TGC TCA TGT GGA GCA CAG ATC ACC GGA 1440  
Asp Gly Ile Met Gln Thr Thr Cys Ser Cys Gly Ala Gln Ile Thr Gly  
465 470 475 480

CAT GTC AAA AAC GGT TCC ATG AGG ATC GTT GGG CCT AAG ACC TGT AGT 1488  
His Val Lys Asn Gly Ser Met Arg Ile Val Gly Pro Lys Thr Cys Ser  
485 490 495

AAC ATG TGG CAT GGA ACA TTC CCC ATC AAC GCA TAC ACC ACG GGC CCC 1536  
Asn Met Trp His Gly Thr Phe Pro Ile Asn Ala Tyr Thr Thr Gly Pro  
500 505 510

TGC ACG CCC TCC CCA GCG CCA AAC TAT TCC AGG GCG CTG TGG CGG GTG 1584  
Cys Thr Pro Ser Pro Ala Pro Asn Tyr Ser Arg Ala Leu Trp Arg Val  
515 520 525

GCT GCT GAG GAG TAC GTG GAG GTT ACG CGG GTG GGG GAT TTC CAC TAC 1632  
Ala Ala Glu Glu Tyr Val Glu Val Thr Arg Val Gly Asp Phe His Tyr  
530 535 540

GTG ACG AGC ATG ACC ACT GAC AAC GTA AAA TGC CCG TGC CAG GTT CCA 1680  
Val Thr Ser Met Thr Asp Asn Val Lys Cys Pro Cys Gln Val Pro  
545 550 555 560

GCC CCC GAA TTC TTC ACA GAA GTG GAT GGG GTG CGG CTG CAC AGG TAC 1728  
Ala Pro Glu Phe Phe Thr Glu Val Asp Gly Val Arg Leu His Arg Tyr  
565 570 575

GCT CCG GCG TGC AAA CCT CTC CTA CGG GAG GTC ACA TTC CAG GTC 1776  
Ala Pro Ala Cys Lys Pro Leu Leu Arg Glu Glu Val Thr Phe Gln Val  
580 585 590

GGG CTC AAC CAA TAC CTG GTT GGG TCG CAG CTC CCA TGC GAG CCC GAA 1824  
Gly Leu Asn Gln Tyr Leu Val Gly Ser Gln Leu Pro Cys Glu Pro Glu  
595 600 605

CCG GAT GTA GCA GTG CTC ACT TCC ATG CTC ACC GAC CCC TCC CAC ATC 1872  
Pro Asp Val Ala Val Leu Thr Ser Met Leu Thr Asp Pro Ser His Ile  
610 615 620

ACA GCA GAG ACG GCT AAG CGC AGG CTG GCC AGG GGG TCT CCC CCC TCC 1920  
Thr Ala Glu Thr Ala Lys Arg Arg Leu Ala Arg Gly Ser Pro Pro Ser  
625 630 635 640

TTG GCC AGC TCT TCA GCT AGC CAG TTG TCT GCG CCT TCC TCG AAG GCG 1968  
Leu Ala Ser Ser Ser Ala Ser Gln Leu Ser Ala Pro Ser Ser Lys Ala  
645 650 655

ACA TAC ATT ACC CAA AAT GAC TTC CCA GAC GCT GAC CTC ATC GAG GCC 2016  
Thr Tyr Ile Thr Gln Asn Asp Phe Pro Asp Ala Asp Leu Ile Glu Ala  
660 665 670

AAC CTC CTG TGG CGG CAT GAG ATG GGC GGG GAC ATT ACC CGC GTG GAG 2064  
Asn Leu Leu Trp Arg His Glu Met Gly Gly Asp Ile Thr Arg Val Glu  
675 680 685

TCA GAG AAC AAG GTA GTA ATC CTG GAC TCT TTC GAC CCG CTC CGA GCG 2112  
Ser Glu Asn Lys Val Val Ile Leu Asp Ser Phe Asp Pro Leu Arg Ala  
690 695 700

GAG GAG GAT GAG CGG GAA GTG TCC GTC CCG GCG GAG ATC CTG CGG AAA 2160  
Glu Glu Asp Glu Arg Glu Val Ser Val Pro Ala Glu Ile Leu Arg Lys  
705 710 715 720

TCC AAG AAA TTC CCA CCA GCG ATG CCC GCA TGG GCA CGC CCG GAT TAC 2208  
Ser Lys Lys Phe Pro Pro Ala Met Pro Ala Trp Ala Arg Pro Asp Tyr  
725 730 735

AAC CCT CCG CTG CTG GAG TCC TGG AAG GCC CCG GAC TAC GTC CCT CCA 2256  
Asn Pro Pro Leu Leu Glu Ser Trp Lys Ala Pro Asp Tyr Val Pro Pro  
740 745 750

GTG GTA CAT GGG TGC CCA CTG CCA CCT ACT AAG ACC CCT CCT ATA CCA 2304  
Val Val His Gly Cys Pro Leu Pro Pro Thr Lys Thr Pro Pro Ile Pro  
755 760 765

CCT CCA CGG AGG AAG AGG ACA GTT GTT CTG ACA GAA TCC ACC GTG TCT 2352  
Pro Pro Arg Arg Lys Arg Thr Val Val Leu Thr Glu Ser Thr Val Ser  
770 775 780

TCT	GCC	CTG	GCG	GAG	CTT	GCC	ACA	AAG	GCT	TTC	GGT	AGC	TCC	GAA	CCG	2400
Ser	Ala	Leu	Ala	Glu	Leu	Ala	Thr	Lys	Ala	Phe	Gly	Ser	Ser	Glu	Pro	
785															800	
TCG	GCC	GTC	GAC	AGC	GGC	ACG	GCA	ACC	GCC	CCT	CCT	GAC	CAA	CCC	TCC	2448
Ser	Ala	Val	Asp	Ser	Gly	Thr	Ala	Thr	Ala	Pro	Pro	Asp	Gln	Pro	Ser	
															815	
GAC	GAC	GGC	GGA	GCA	GGA	TCT	GAC	GTT	GAG	TCG	TAT	TCC	TCC	ATG	CCC	2496
Asp	Asp	Gly	Gly	Ala	Gly	Ser	Asp	Val	Glu	Ser	Tyr	Ser	Ser	Met	Pro	
															830	
CCC	CTT	GAG	GGG	GAG	CCG	GGG	GAC	CCC	GAT	CTC	AGC	GAC	GGG	TCT	TGG	2544
Pro	Leu	Glu	Gly	Glu	Pro	Gly	Asp	Pro	Asp	Leu	Ser	Asp	Gly	Ser	Trp	
															845	
TCT	ACC	GTG	ACT	GAG	GAG	GCC	GGT	GAG	GAC	GTC	GTC	TGC	TGC	TCG	ATG	2592
Ser	Thr	Val	Ser	Glu	Glu	Ala	Gly	Glu	Asp	Val	Val	Cys	Cys	Ser	Met	
															860	
TCC	TAC	ACA	TGG	ACA	GGC	GCT	CTG	ATC	ACG	CCA	TGC	GCT	GCG	GAG	GAA	2640
Ser	Tyr	Thr	Trp	Thr	Gly	Ala	Leu	Ile	Thr	Pro	Cys	Ala	Ala	Glu	Glu	
															880	
AGC	AAG	CTG	CCC	ATC	AAC	GCG	TTG	AGC	AAC	TCT	TTG	CTG	CGT	CAC	CAC	2688
Ser	Lys	Leu	Pro	Ile	Asn	Ala	Leu	Ser	Asn	Ser	Leu	Leu	Arg	His	His	
															895	
AAC	ATG	GTC	TAC	GCT	ACC	ACA	TCC	CGC	AGC	GCA	AGC	CAG	CGG	CAG	AAG	2736
Asn	Met	Val	Tyr	Ala	Thr	Thr	Ser	Arg	Ser	Ala	Ser	Gln	Arg	Gln	Lys	
															910	
AAG	GTC	ACC	TTT	GAC	AGA	CTG	CAA	ATC	CTG	GAC	GAT	CAC	TAC	CAG	GAC	2784
Lys	Val	Thr	Phe	Asp	Arg	Leu	Gln	Ile	Leu	Asp	Asp	His	Tyr	Gln	Asp	
															925	

GTG CTC AAG GAG ATG AAG GCG AAG GCG TCC ACA GTT AAG GCT AAG CTT	2832	
Val Leu Lys Glu Met Lys Ala Lys Ala Ser Thr Val Lys Ala Lys Leu		
930	935	940
CTA TCA GTA GAG GAA GCC TGC AAG CTG ACG CCC CCA CAT TCG GCC AAA	2880	
Leu Ser Val Glu Glu Ala Cys Lys Leu Thr Pro Pro His Ser Ala Lys		
945	950	955
960		
TCT AAA TTT GGC TAT GGG GCA AAG GAC GTC CGG AAC CTA TCC AGC AAG	2928	
Ser Lys Phe Gly Tyr Gly Ala Lys Asp Val Arg Asn Leu Ser Ser Lys		
965	970	975
GCC ATT AAC CAC ATC CGC TCC GTG TGG GAG GAC TTG TTG GAA GAC ACT	2976	
Ala Ile Asn His Ile Arg Ser Val Trp Glu Asp Leu Leu Glu Asp Thr		
980	985	990
GAA ACA CCA ATT GAC ACC ACC ATC ATG GCA AAA AAT GAG GTT TTC TGC	3024	
Glu Thr Pro Ile Asp Thr Thr Ile Met Ala Lys Asn Glu Val Phe Cys		
995	1000	1005
GTC CAA CCA GAG AGA GGA GGC CGC AAG CCA GCT CGC CTT ATC GTG TTC	3072	
Val Gln Pro Glu Arg Gly Arg Lys Pro Ala Arg Leu Ile Val Phe		
1010	1015	1020
CCA GAC TTG GGG GTC CGT GTG TGC GAG AAA ATG GCC CTC TAT GAC GTG	3120	
Pro Asp Leu Gly Val Arg Val Cys Glu Lys Met Ala Leu Tyr Asp Val		
1025	1030	1035
1040		
GTC TCC ACC CTC CCT CAG GCT GTG ATG GGC TCC TCG TAC GGA TTC CAG	3168	
Val Ser Thr Leu Pro Gln Ala Val Met Gly Ser Ser Tyr Gly Phe Gln		
1045	1050	1055
TAT TCT CCT GGA CAG CGG GTC GAG TTC CTG GTG AAC GCC TGG AAA TCA	3216	
Tyr Ser Pro Gly Gln Arg Val Glu Phe Leu Val Asn Ala Trp Lys Ser		
1060	1065	1070

AAG AAG ACC CCT ATG GGC TTT GCA TAT GAC ACC CGC TGT TTT GAC TCA 3264  
Lys Lys Thr Pro Met Gly Phe Ala Tyr Asp Thr Arg Cys Phe Asp Ser  
1075 1080 1085

ACA GTC ACT GAG AAT GAC ATC CGT GTA GAG GAG TCA ATT TAT CAA TGT 3312  
Thr Val Thr Glu Asn Asp Ile Arg Val Glu Glu Ser Ile Tyr Gln Cys  
1090 1095 1100

TGT GAC TTG GCC CCC GAA GCC AGA CAG GCC ATA AGG TCG CTC ACA GAG 3360  
Cys Asp Leu Ala Pro Glu Ala Arg Gln Ala Ile Arg Ser Leu Thr Glu  
1105 1110 1115 1120

CGG CTT TAT ATC GGG GGT CCC CTG ACT AAT TCA AAA GGG CAG AAC TGC 3408  
Arg Leu Tyr Ile Gly Gly Pro Leu Thr Asn Ser Lys Gly Gln Asn Cys  
1125 1130 1135

GGC TAT CGC CGG TGC CGC GCG AGC GGC GTG CTG ACG ACT AGC TGC GGT 3456  
Gly Tyr Arg Arg Cys Arg Ala Ser Gly Val Leu Thr Thr Ser Cys Gly  
1140 1145 1150

AAT ACC CTC ACA TGT TAC TTG AAG GCC TCT GCA GCC TGT CGA GCT GCA 3504  
Asn Thr Leu Thr Cys Tyr Leu Lys Ala Ser Ala Ala Cys Arg Ala Ala  
1155 1160 1165

AAG CTC CAG GAC TGC ACG ATG CTC GTG TGC GGA GAC GGC CTT GTC GTT 3552  
Lys Leu Gln Asp Cys Thr Met Leu Val Cys Gly Asp Asp Leu Val Val  
1170 1175 1180

ATC TGT GAG AGC GCG GGA ACC CAG GAG GAC GCG GCG AGC CTA CGA GTC 3600  
Ile Cys Glu Ser Ala Gly Thr Gln Glu Asp Ala Ala Ser Leu Arg Val  
1185 1190 1195 1200

TTC ACG GAG GCT ATG ACT AGG TAC TCT GCC CCC CCC GGG GAC CCG CCC 3648  
Phe Thr Glu Ala Met Thr Arg Tyr Ser Ala Pro Pro Gly Asp Pro Pro  
1205 1210 1215

CAA CCA GAA TAC GAC CTG GAG TTG ATA ACA TCA TGC TCC TCC AAT GTG 3696  
Gln Pro Glu Tyr Asp Leu Glu Leu Ile Thr Ser Cys Ser Ser Asn Val

1220

1225

1230

TCG GTC GCG CAC GAT GCA TCT GGC AAA AGG GTA TAC TAC CTC ACC CGT 3744  
Ser Val Ala His Asp Ala Ser Gly Lys Arg Val Tyr Tyr Leu Thr Arg

1235

1240

1245

GAC CCG

3750

Asp Pro

1250

DRAFT - DO NOT CITE

SEQ ID NO:23

SEQUENCE TYPE:Nucleotide

SEQUENCE LENGTH:23 BASES

STRANDEDNESS:single

TOPOLOGY:linear

MOLECULE TYPE:synthetic DNA

ORIGINAL SOURCE ORGANISM:baculovirus *Autographa californica* Nuclear Polyhedrosis virus (AcNPV)

IMMEDIATE EXPERIMENTAL SOURCE:Oligonucleotide synthesiser; oligo d24

FEATURES:

from 1 to 23 bases homologous to portion of AcNPV polyhedrin gene downstream of the BamH1 cloning site in pAc360 and similar vectors

PROPERTIES:primes DNA synthesis from baculovirus transfer vector sequences which flank DNA inserted at the BamH1 site.

CGGGTTTAAC ATTACGGATT TCC

23

SEQ ID NO:24

SEQUENCE TYPE:Nucleotide

SEQUENCE LENGTH:31 BASES

STRANDEDNESS:single

TOPOLOGY:linear

MOLECULE TYPE:synthetic DNA

ORIGINAL SOURCE ORGANISM:baculovirus Autographa californica Nuclear Polyhedrosis virus (AcNPV)

IMMEDIATE EXPERIMENTAL SOURCE:Oligonucleotide synthesiser; oligo d126

FEATURES:

from 1 to 31 bases homologous to the upstream junction sequences produced when cDNA amplified by d75 (SEQ ID 5) is cloned into the BamH1 cloning site in pAc360 and similar vectors; mismatches at bases 13 and 14 introduce a PstI site

from 1 to 10 bases homologous to region of BamH1 site in pAc360 and similar vectors

from 4 to 9 bases BamH1 site

from 12 to 17 bases PstI site

PROPERTIES:primes DNA synthesis at the junction of baculovirus transfer vector sequences and sequences previously amplified by oligo d75; introduces a PstI recognition site for subsequent cloning work

TAAGGATCCC CCT GCA GTA TCG GCG GAA TTC

31

Ser Ala Val Ser Ala Glu Phe

SEQ ID NO:25

SEQUENCE TYPE:Nucleotide

SEQUENCE LENGTH:45 BASES

STRANDEDNESS:single

TOPOLOGY:linear

MOLECULE TYPE:synthetic DNA

ORIGINAL SOURCE ORGANISM:N/A

IMMEDIATE EXPERIMENTAL SOURCE:Oligonucleotide synthesiser; oligo d132

FEATURES:

from 5 to 10 bases PstI recognition site

from 13 to 27 bases linker coding for five Lys residues

from 28 to 45 bases homologous to bases 4 to 21 of BR11 (SEQ ID 7)

PROPERTIES:primes DNA synthesis at the 5' end of BR11 and introduces a synthetic sequence which codes for five lysines as well as a PstI recognition site for subsequent cloning work

CTGCCTGCA GTA AAG AAG AAG AAG AAG AAA ACC AAA CGT AAC ACC A        45

Val Lys Lys Lys Lys Lys Lys Thr Lys Arg Asn Leu

5

10